Supplement: Structured Uncertainty Prediction Networks

Garoe Dorta [1,2] Sara Vicente [2] Lourdes Agapito [3] Neill D. F. Campbell [1] Ivor Simpson [2]
[1]University of Bath [2]Anthropics Technology Ltd. [3]University College London
[1]{g.dorta.perez,n.campbell}@bath.ac.uk [2]{sara,ivor}@anthropics.com [3]l.agapito@cs.ucl.ac.uk

A. Network architectures

All the models were trained with a batch size of 64. The exp block in all the architectures removes the log in the diagonal values of the Cholesky matrix that is being estimated.

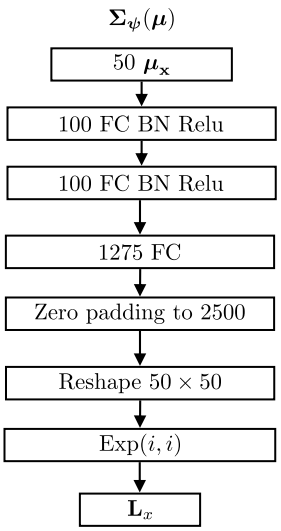


Figure 1: Covariance estimation network architecture for the splines dataset.

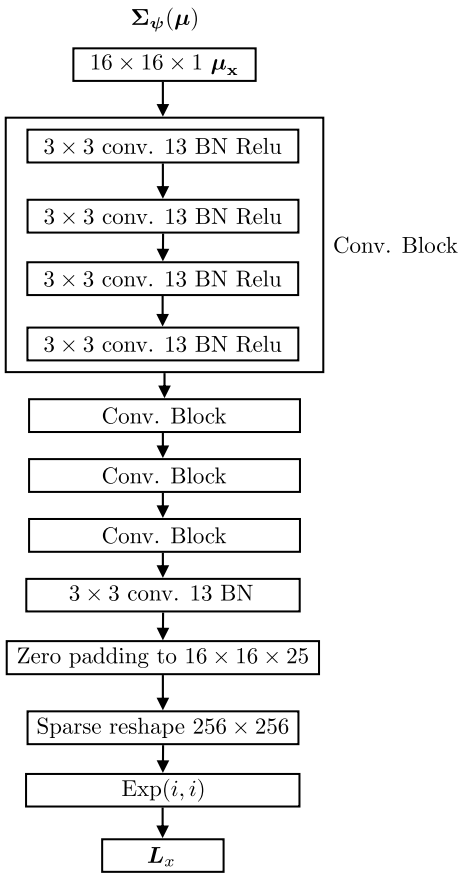


Figure 2: Covariance estimation network architecture for the ellipses dataset.

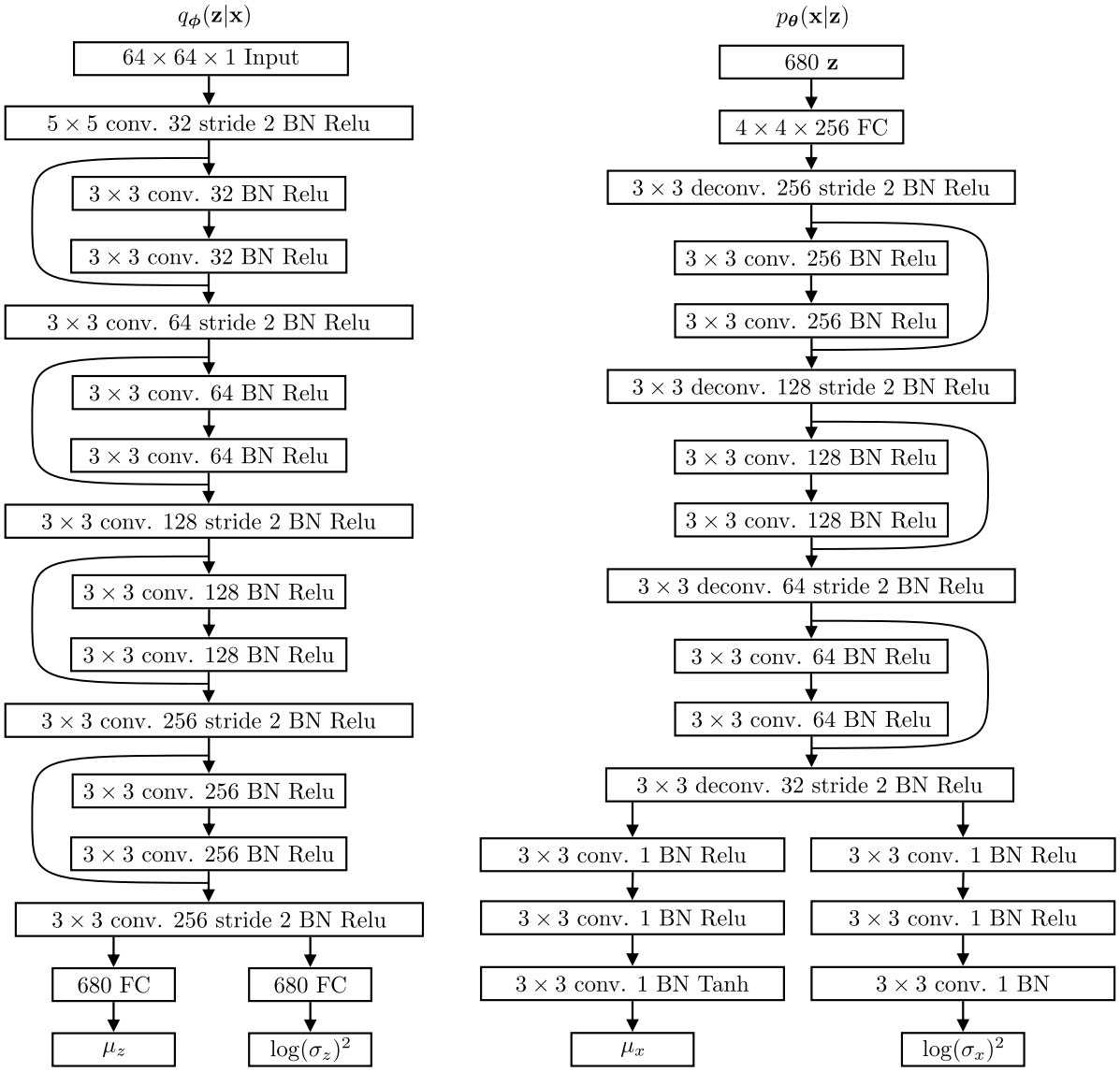


Figure 3: VAE architecture for the CelebA dataset.

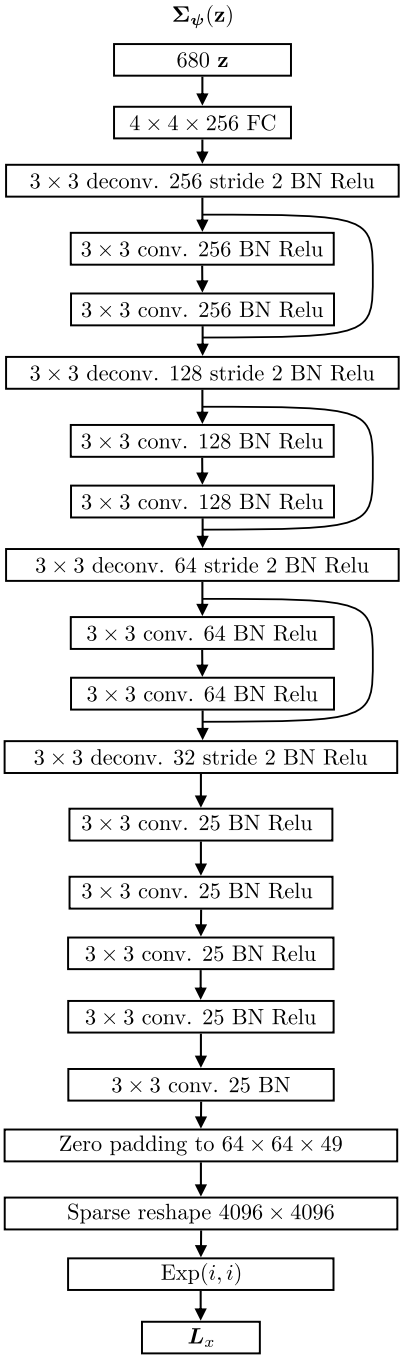


Figure 4: Covariance estimation network architecture for the CelebA dataset.

B. CIFAR10 dataset

In this dataset our models uses a patch size of 5×5, and an architecture adapted from the CelebA model for 32×32 images. Quantitatively, a VAE with diagonal covariance achieved a marginal negative log likelihood of -1026 ± 462 and our model of -1333 ± 477, where the likelihood was evaluated with 500 $\mathbf{z}$ samples per image.

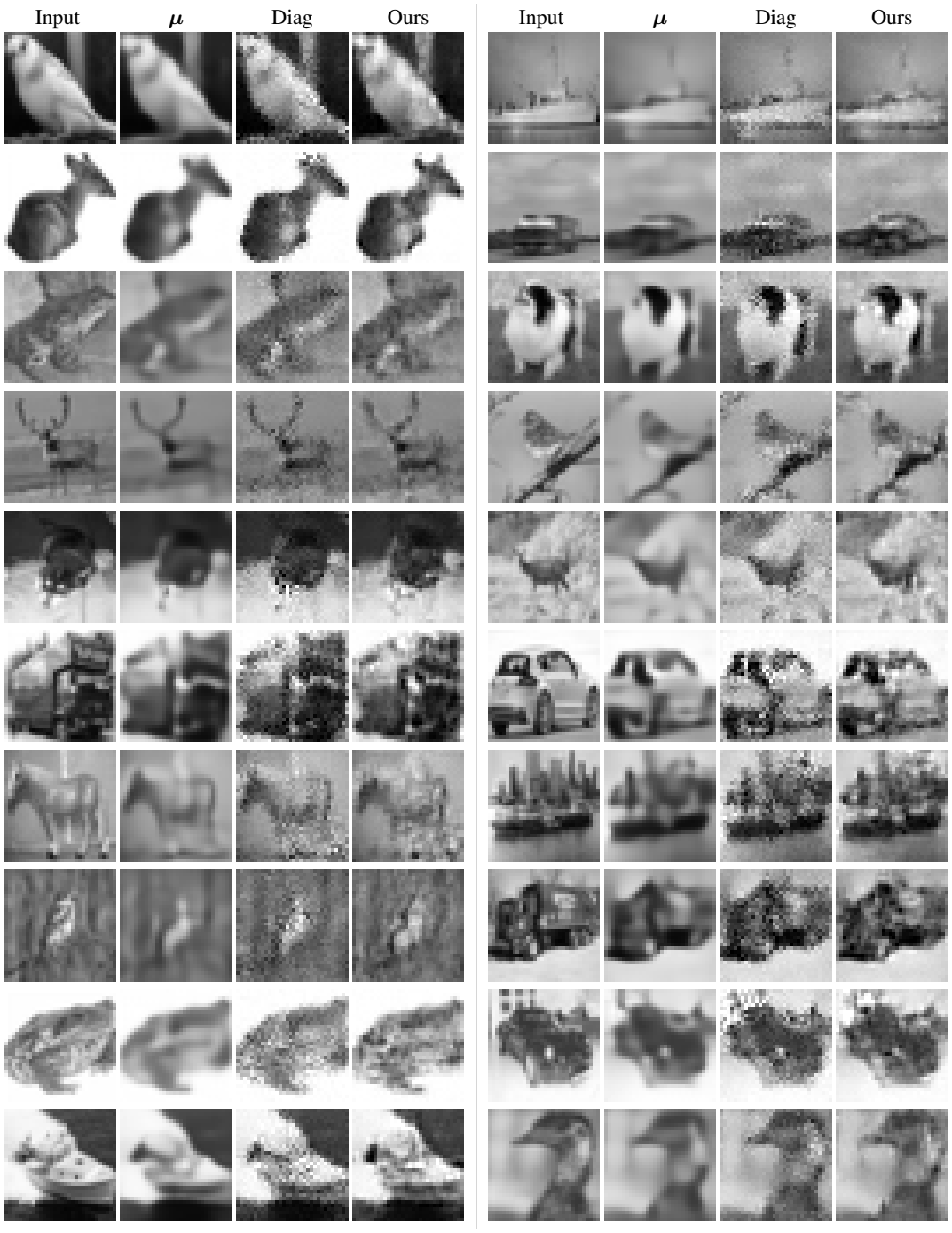
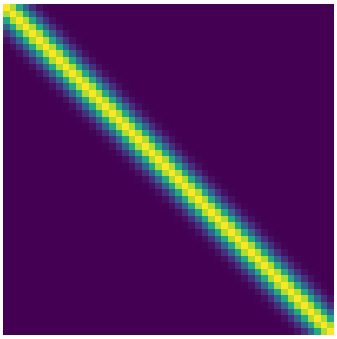


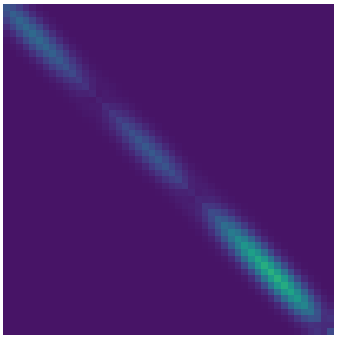
Figure 5: Comparison of image reconstructions on the CIFAR10 dataset, where $\boldsymbol{\mu}$ is a reconstruction from a VAE, Diag. is $\boldsymbol{\mu} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a sample from a diagonal covariance matrix, and in our method $\boldsymbol{\epsilon}$ is a sample from a dense covariance matrix.

C. Splines dataset

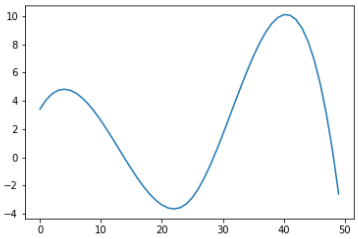
In the splines dataset, for each example, $\boldsymbol{\mu}$ is a cubic spline interpolation of five random points sampled from a Gaussian distribution, as shown in Fig. 6c. Given a predefined prototype covariance matrix (shown in Fig. 6a), the covariance matrix $\boldsymbol{\Sigma}$ for a particular example is constructed by scaling this prototype covariance by the absolute value of $\boldsymbol{\mu}$, as shown in Fig. 6b. Finally, a random sample is drawn from the covariance matrix and added to the mean $\boldsymbol{\mu}$, which generates the final example $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$, as shown in Fig. 6d.



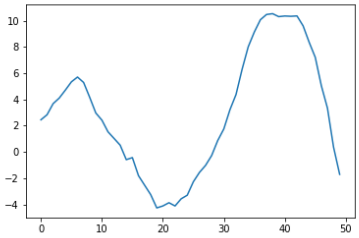
(a) Prototype covariance



(b) Example covariance



(c) Example mean $\boldsymbol{\mu}$



(d) Example mean $\boldsymbol{\mu}$ plus sample from covariance

Figure 6: The splines dataset is synthesized using a prototype covariance matrix (a), that is transformed to generate a number of example covariances (b), where each transformation is function of a different spline (c). Each example (d) in the dataset is constructed by taking a single sample from a covariance (b) and adding it to the corresponding spline (c).

D. CelebA

Figure 7: Denoising experiment, left column: original image without noise, second column: image with added noise, third column: denoising autoencoder (DAE) result, fourth column: our result, fifth column: VAE reconstruction from the noisy input, sixth column: difference between the VAE reconstruction and the noisy input, right column: the difference projected on $\hat{\mathbf{\Sigma}}$, the matrix constructed with 1000 eigenvectors of $\mathbf{\Sigma}$ with the largest eigenvalues. Our result is the sum of the projected difference and the VAE reconstruction. Our model is able to recover fine details that are lost with the DAE approach.

Figure 8: Denoising experiment, left column: original image without noise, second column: image with added noise, third column: denoising autoencoder result, and right column: our result. Our model is able to recover fine details with its structure residual prediction.

Figure 9: Comparison of image reconstructions for the different models. The best of 100 and 1000 samples from $\mathbf{\Sigma}$ measured using the MSE to the input are shown. The AE and VAE both generate over-smoothed images. For both the AE and VAE, our model adds plausible high-frequencies from a single sample drawn from the predicted uncertainty distribution.

Figure 10: Images generated by decoding samples from the prior distribution on the latent space of a β-VAE with added residuals from our model, where $\beta = 5$.

Figure 11: Images generated by decoding samples from the prior distribution on the latent space of a VAE.

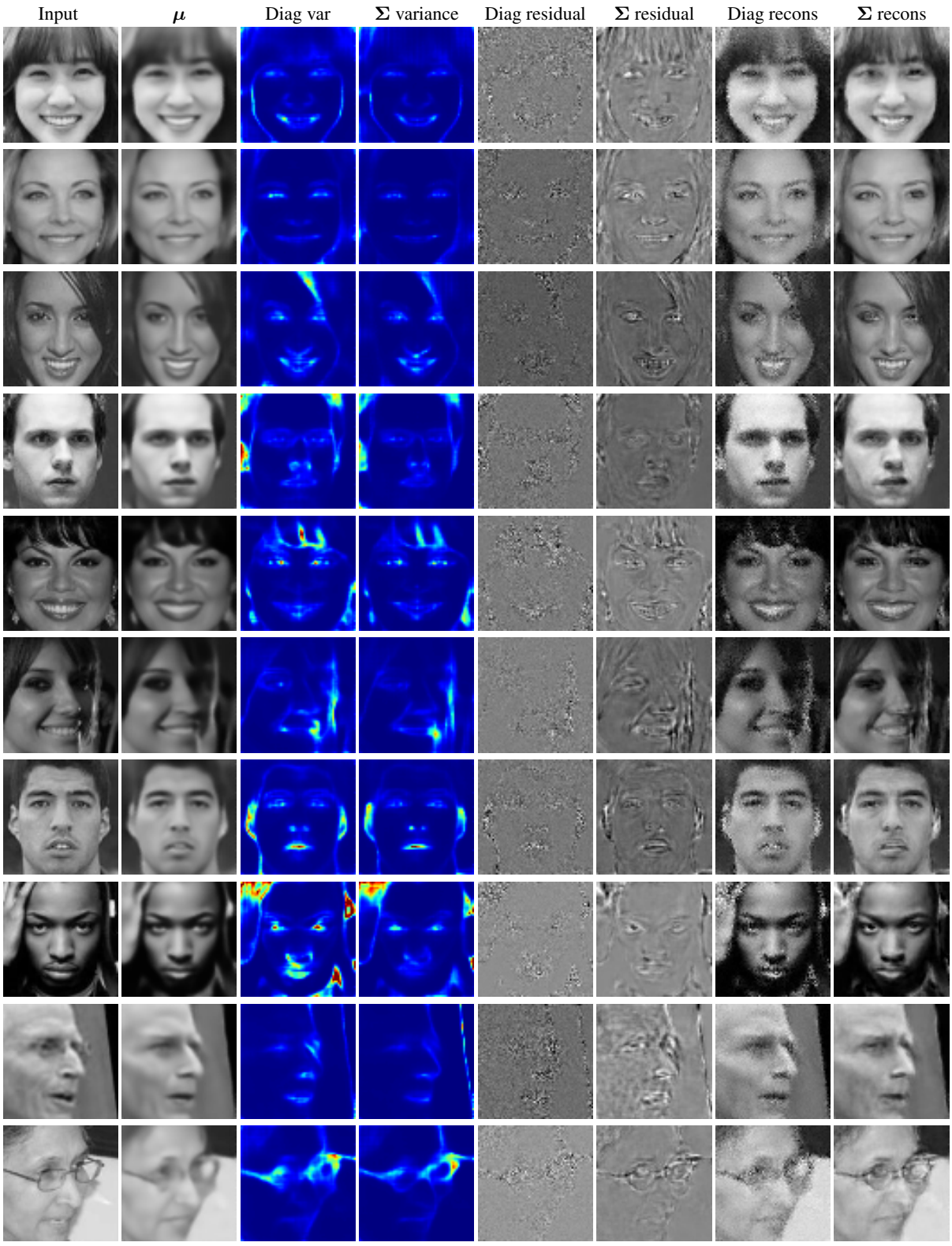


Figure 12: Variance maps for different inputs, taking a single sample from the estimated residual distribution. The diagonal noise estimation model mistakenly identifies teeth or skin wrinkles as variance, whereas the covariance model properly identifies them as regions with high covariance, yet low variance.

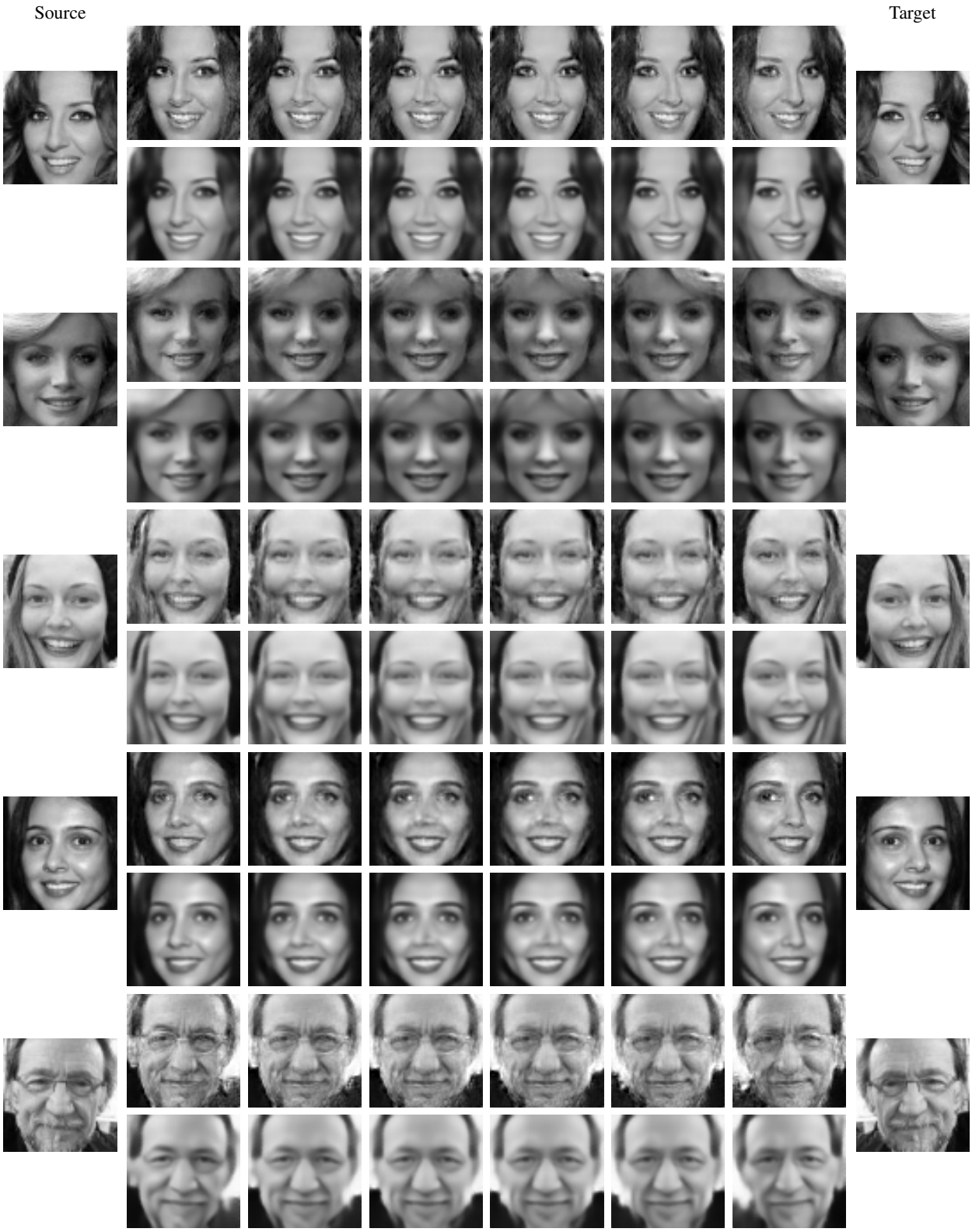


Figure 13: Samples drawn with our model while interpolating on the latent space, from the source to target. Using a fixed noise vector $\mathbf{u}$, a single sample is drawn from our model as $\mathbf{x} = \boldsymbol{\mu} + \mathbf{M}\mathbf{u}$, where the covariance $\boldsymbol{\Sigma} = \mathbf{M}\mathbf{M}^T$ is estimated from each $\mathbf{z}$. The latent values $\mathbf{z}$ are interpolated using polar interpolation. For every pair of rows, the first contains the result from our model, and the second contains the reconstruction $\boldsymbol{\mu}(\mathbf{z})$ from the VAE.